

Figure 1

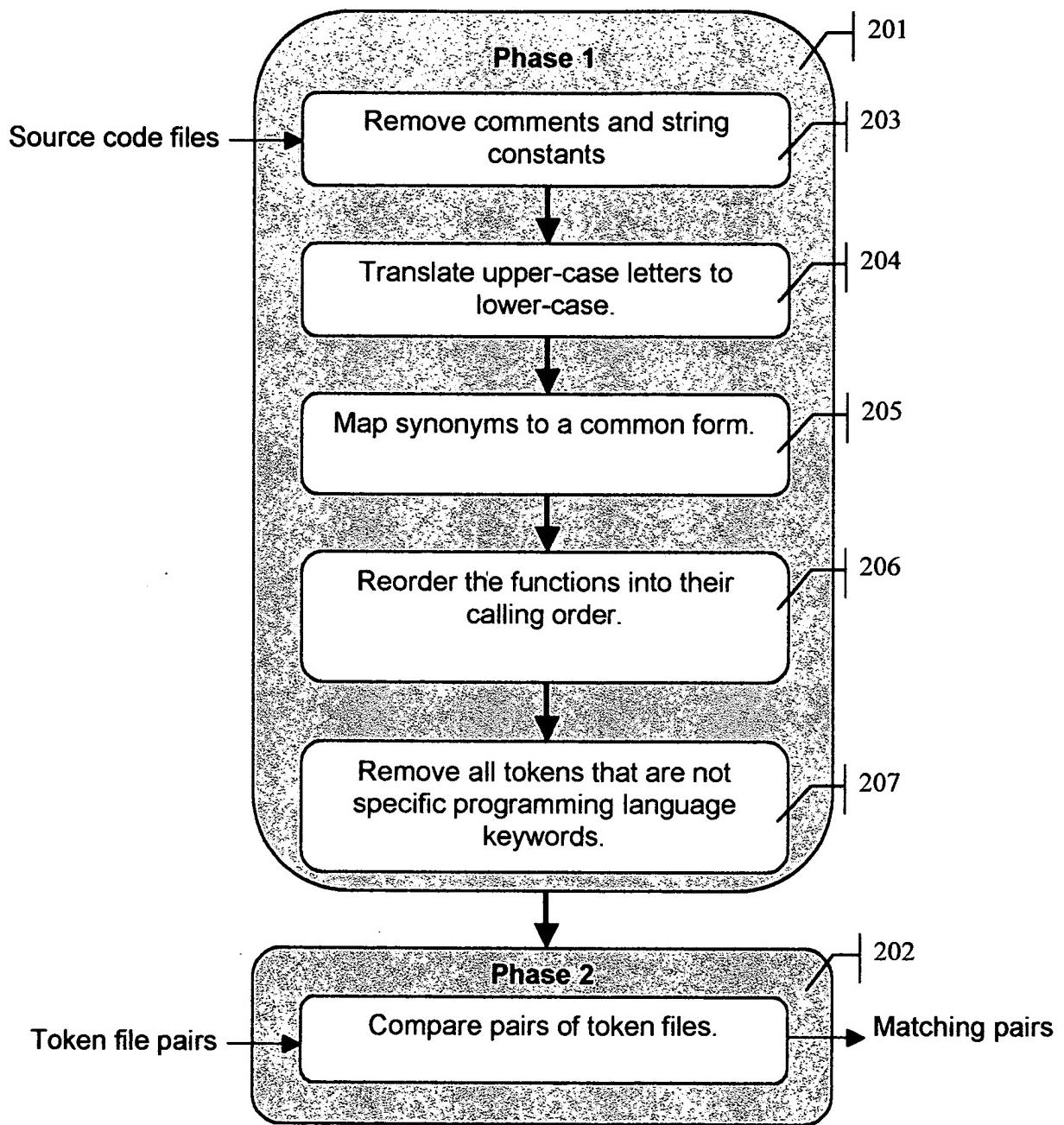
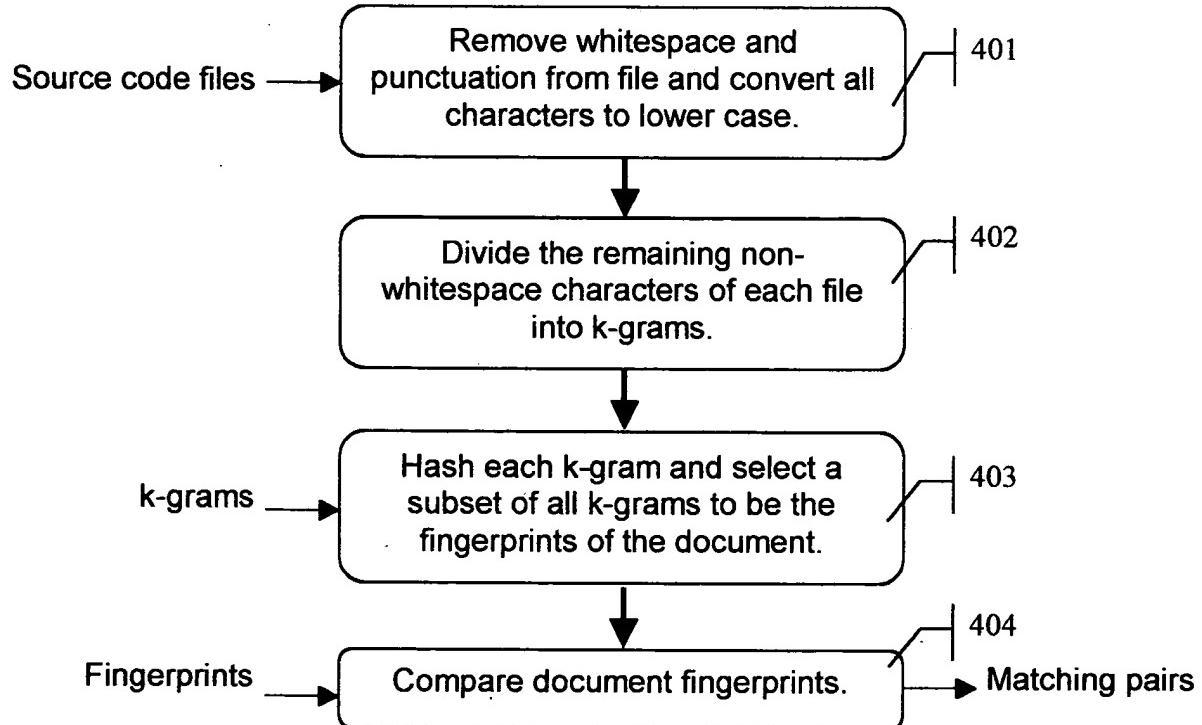
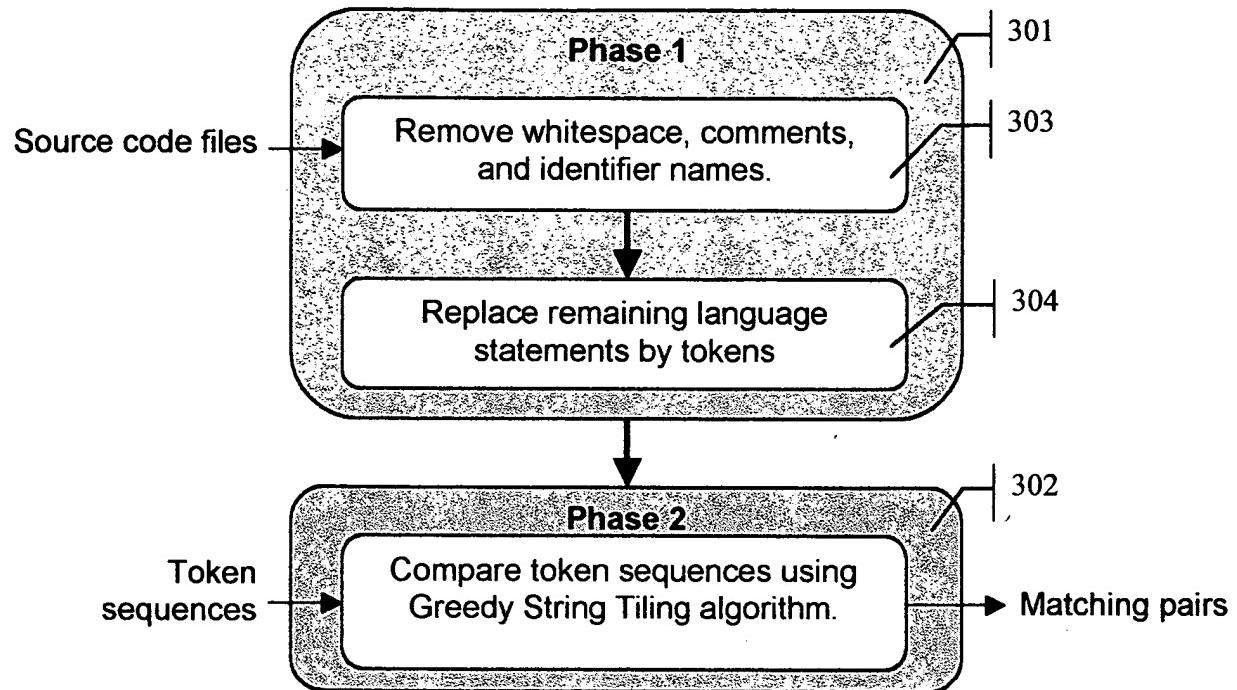


Figure 2



She loves you yeah, yeah, yeah. | 501
(a) Some text.

shelo helov elove loves ovesy vesyo esyou syou youye
 ouyea uyeah yeahy eahye ahyea hyeah yeahy eahye ahyea
 hyeah

(b) The sequence of 5-grams derived from the text. | 502

77 72 42 17 98 50 23 55 6 66 34 24 39 11 84 24 39 11 84

(c) A hypothetical sequence of hashes of the 5-gram. | 503

72 24 84 24 84

(d) The fingerprints – selecting only those hashes that are 0 mod 4. | 504

Figure 5

```

/* ----- begin routine ----- */
void fdiv(
    char      *fname,        // file name
    char *path)           // path
{
    int Index1, j;          | 601
    while (1)
        j =      strlen(fname);
    /* find the file extension */ | 602

```

(a) C source code snippet for file 1.

```

SourceLines1[0] = ""
SourceLines1[1] = "void fdiv"
SourceLines1[2] = "char fname"
SourceLines1[3] = "char path"
SourceLines1[4] = ""
SourceLines1[5] = "int Index1 j"
SourceLines1[6] = ""
SourceLines1[7] = "while 1"
SourceLines1[8] = "j strlen fname"
SourceLines1[9] = ""
CommentLines1[0] = "----- begin routine -----"
CommentLines1[1] = ""
CommentLines1[2] = "file name"
CommentLines1[3] = "path"
CommentLines1[4] = ""
CommentLines1[5] = ""
CommentLines1[6] = ""
CommentLines1[7] = ""
CommentLines1[8] = ""
CommentLines1[9] = "find the file extension"

```

(b) Source code and comment line arrays for file 1.

```

Word1[0] = "fdiv"
Word1[1] = "fname"
Word1[2] = "path"
Word1[3] = "Index1" | 603

```

(c) Array of unique identifiers (non-keywords) in file 1.

Figure 6

```

Word1[0] = "abc"           Word2[0] = "Aabc"
Word1[1] = "abc1"          Word2[1] = "aBc"
Word1[2] = "abc123"        Word2[2] = "abc111111"
Word1[3] = "abcdef"        Word2[3] = "abcXXXyz"
Word1[4] = "pdq"           Word2[4] = "i"
Word1[5] = "xxx"           Word2[5] = "j"
Word1[6] = "xyz"           Word2[6] = "pdq"
Word1[7] = "yyy"           Word2[7] = "X"

```

(a) Non-keyword words in files 1 and 2.

```

PartialWord[0] = "abc"
PartialWord[1] = "abc1"
PartialWord[2] = "xxx"
PartialWord[3] = "xyz"

```

(b) Matching partial words

701

702

Figure 7

File 1	File 2
1 /* ---- begin routine ---- */	1 /* find the file extension */
2 void fdiv(2 void file_divide(
3 char *fname, // file name	3 char *fname,
4 char *path) /* path */	4 char *path)
5 {	5 {
6 int Index1, j;	6 int i, j;
7 while (1)	7 while (1) // loop here
8 j = strlen(fname);	8 j = strlen(fname);
9 }	9 }
10 // find the file extension	10 }

(a) Two files.

3/3
4/4
9/8

(b) Matching source lines in file1/file2

801

802

Figure 8

File 1	File 2
1 /* ---- begin routine ---- */	1 /* find the file extension */
2 void fdiv(2 void file_divide(
3 char *fname, // file name	3 char *fname,
4 char *path) /* path */	4 char *path) // path
5 {	5 {
6 int Index1, j;	6 int i, j; /* ---- begin routine ---- */
7	7 while (1) // loop here
8 while (1)	8 j = strlen(fname);
9 j = strlen(fname);	9
10 // find the file extension	10 switch (x)
11 if (x == 5) {	11 {

(a) Two files.

1/6
4/4
10/1

(b) Matching comment lines in file1/file2

901

902

Figure 9

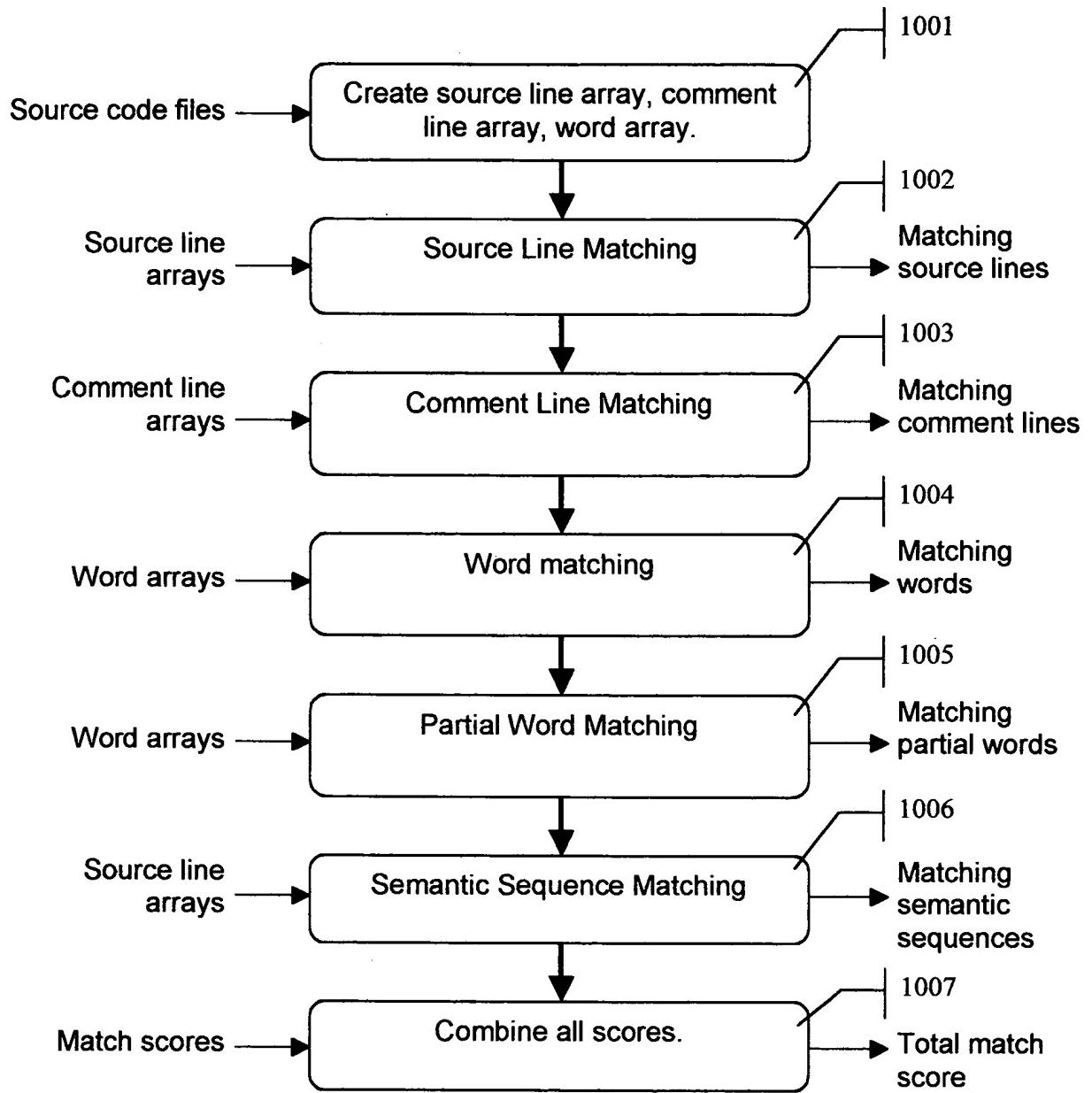
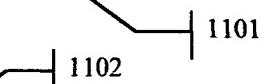


Figure 10

Comparing files in folder D:\CodeMatch\Code Development\test\C test 2\files 1
To files in folder D:\CodeMatch\Code Development\test\C test 2\files 2


D:\CodeMatch\Code Development\test\C test 2\files 1\bpf_dump.c


Match Score Compared To File

<u>2910</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\bpf_dump.c
<u>374</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\W32NReg.c
<u>374</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\test\W32NReg (variable names changed).c
<u>374</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\test\W32NReg (no comments).c

D:\CodeMatch\Code Development\test\C test 2\files 1\bpf_filter.c


Match Score Compared To File

<u>606</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\W32NReg.c
<u>606</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\test\W32NReg (no comments).c
<u>572</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\test\W32NReg (variable names changed).c
<u>398</u>	D:\CodeMatch\Code Development\test\C test 2\files 2\bpf_dump.c

Figure 11

**Comparing file1: D:\CodeMatch\test\C test 2\files 1\bpf_dump.c
To file2: D:\CodeMatch\ C test 2\files 2\test\W32Nreg.c**

Matching source lines:

File1 Line#	File2 Line#	Source line
21	1	#include <windows.h>
22	3	#include <stdio.h>
24	7	#include "WiNDIS.h"

1201

1202

1203

Matching comment lines:

File1 Line#	File2 Line#	Comment line
3	3	* The Regents of the University of California. All rights reserved.
10	5	* Redistribution and use in source and binary forms, with or without

1204

Longest matching semantic sequence:

File1 Line# File2 Line# Number of matching lines

21 1 3

1205

Matching words:

stdio	WiNDIS	windows
-------	--------	---------

1206

Matching partial words:

0x	windows
----	---------

Figure 12